LA-UR-19-25748
Approved for public release; distribution is unlimited.

Title: Measurement Error Variance Estimation in Gamma-Spectroscopy Data using FRAM

Author(s): Burr, Thomas Lee
Sampson, Thomas
Vo, Duc Ta

Intended for: Report

Issued: 2019-06-20

Measurement Error Variance Estimation in Gamma-Spectroscopy Data using FRAM

Tom Burr (CCS-6), Tom Sampson (Sampson Professional Services), Duc Vo (NEN-1)

Abstract

The FRAM gamma-ray isotopic analysis code has recently been applied to four Pu data sets and four U data sets, each with relatively well-characterized standard items, in order to estimate the variances of random errors and of item-specific systematic errors. Assay results are available for Pu238, Pu239, Pu240, Pu241, Pu242 (operator entered, not independently measured), Am241, Specific Power, and Pu240eff. For the four U data sets, results are available for the measurands U234, U235, U236 (operator entered, not independently measured), and U238. For example, in Pu data set 1, there are 33 standard reference Pu items, which have well-known nominal/true values (except for a few items). FRAM was used to make approximately 15 repeat measurements for each of the 33 items, so empirical standard deviations can be calculated. In addition, FRAM internally estimates its random error variance using propagation of variance, as explained in [1]. This paper applies analysis of variance (ANOVA) to separately estimate the repeatability standard deviation and the standard deviation of the item-specific biases δ_S in each of the 8 data sets.

1.Introduction and Background

The FRAM (Fixed-energy Response-function Analysis with Multiple Efficiencies method) gamma-ray isotopic analysis software [1] has been used to analyse gamma-spectroscopy data taken with a high resolution (High purity Germanium, HPGe) detector. FRAM has recently been applied to four Pu data sets and four U data sets. In Pu data set 1, there are 33 standard reference Pu items, which have well-known nominal/true values except in a few cases as noted. For the four Pu data sets, results are available for the measurands Pu238, Pu239, Pu240, Pu241, Pu242, Am241, Specific Power, and Pu240eff. For the four U data sets, results are available for the measurands U234, U235, U236, U238.

Uncertainty quantification (UQ) can be approached from first physical principles ("bottom-up") or empirically ("top-down") [2-4]. This paper presents top-down analyses four Pu data sets and four U data sets collected with a HPGe detector and analysed with FRAM [1]. Analysis of variance (ANOVA) is applied to separately estimate the relative standard deviation (RSD) among repeated measurements of the same item, and the RSD of the item-specific biases [5-7]. All eight data sets can be regarded as being repeated measurements on standards. The nominal ("true") values of each standard "item" is available (based on mass spectrometry).

Data set 1 of 8 (Pu) has 33 items, with approximately 15 repeated measurements per item for a total of 490 measurements. To illustrate, Figure 1 is the 490 measurements of Pu239 plotted as (FRAM-Massspec)/Massspec = $(M\text{-}T)/T$ with the mass spectrometry measurement regarded as the true (T) value for each of the 33 (33= 6 *x* 5 + 3) items in data set 1. The average value of $(M\text{-}T)/T$ for each item is an estimate of the item-specific bias and is plotted as a horizontal line. Note from informal assessment of Figure 1 that the standard deviation (SD) of the item-specific biases (approximately equal to the SD of the between-item means; see Section 2) is similar in magnitude to the repeatability SD. The ANOVA-based estimate of the standard deviation of the item-specific biases δ_S is given in Appendix 1, and the ANOVA procedure is described in Section 2. Figure 2 plots the empirical standard deviation $\hat{\sigma} = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)}$ versus the reference value for each of the 8 measurands. The value of $\hat{\sigma}$ for Pu242 is 0 because the "measured" value is operator-entered using

the reference "true" value. All other $\hat{\sigma}$ values increase with the reference value, indicating that a multiplicative error model is appropriate, except for Pu239, where $\hat{\sigma}$ decreases as the reference value increases. Figure 3 is the same as Figure 2, but plots FRAM's estimated SD (using variance propagation, see [1,2]) versus the reference value. FRAM's estimate of the SD includes some error sources in addition to those from Poisson variation. For example, in data set 1, the average ratio (over all 8 measurands except for Pu242) of the observed (empirical) SD to FRAM's estimated SD is 0.91, and an approximate 95% confidence interval is (0.82, 0.99), indicating that the empirical SD is slightly smaller than FRAM's estimated of the SD. Note that Pu242 is entered by the FRAM operator, so any "errors" are due to rounding, and so the Pu242 results are included here only for completeness.

Figure 4 is the estimation error (FRAM estimate – nominal value) versus the reference value. Figure 5 is the same as Figure 4, except it plots the relative estimation error (estimated relative bias) and indicates the sample identity using the integers 1 to 33. The item-specific bias tends to have larger magnitude as the reference value increases, except for Pu239 (ignore Pu242 here because Pu242 is entered by the operator). Figure 6 plots the estimated ratio $\frac{\delta_S}{\delta_R}$ for each of the 4 Pu data sets with 8 measurands. Figure 7 is the same as Figure 6, but for each of the 4 U data sets with 4 measurands.

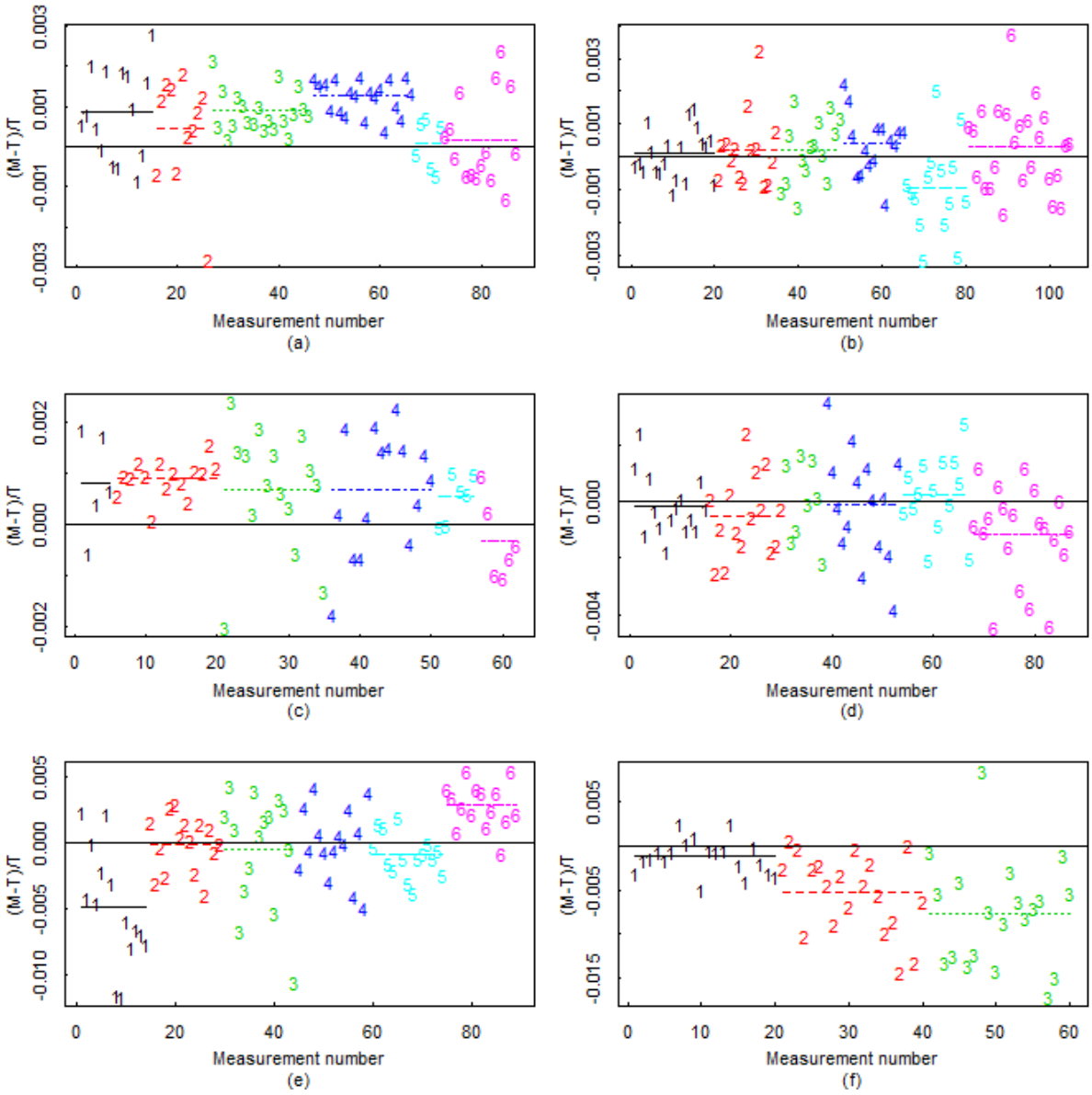

Figure 1. The 490 measurements of Pu239 plotted as (FRAM-Massspec)/Massspec = (*M*-*T*)/*T* with massspec regarded as the true (T) value for each of the 33 (33= 6 *x* 5 + 3) items in data set 1. The average value of (*M*-*T*)/*T* for each item is an estimate of the item-specific bias and is plotted as a horizontal line. The SD of the item-specific biases is similar in magnitude to the repeatability SD.

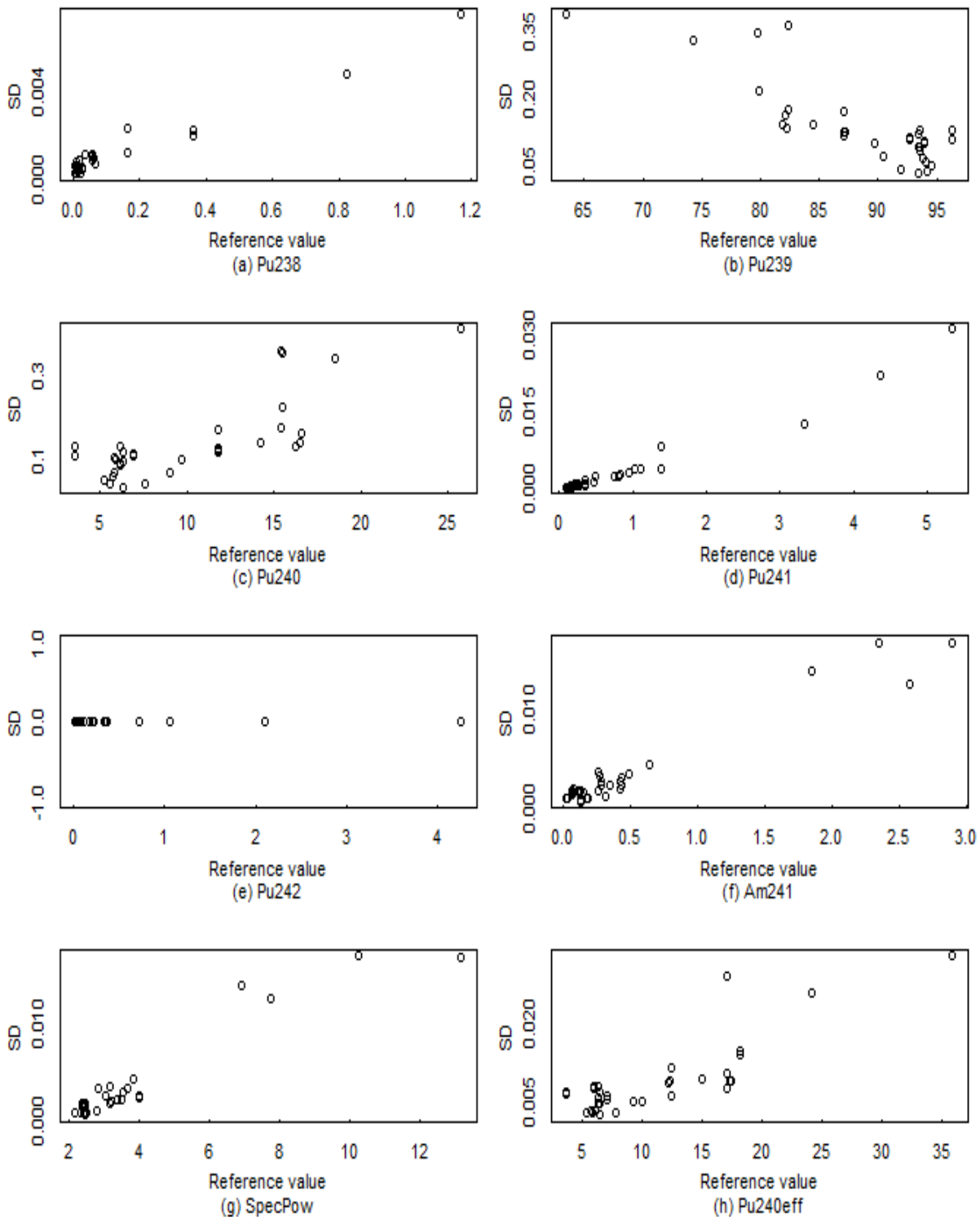

Figure 2. The empirical standard deviation (SD) vs. the reference value for each of the 8 measurand in data set 1.

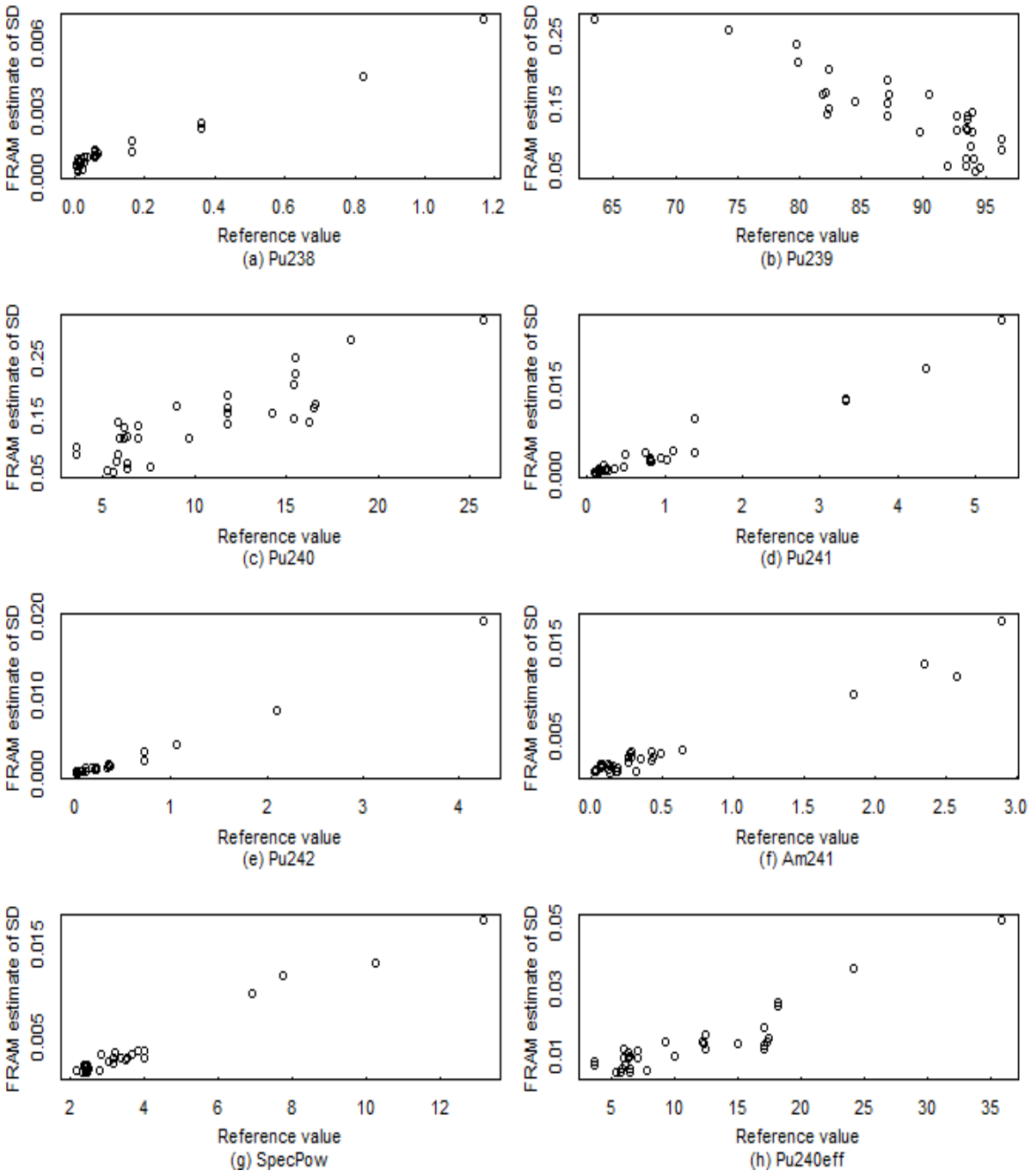

Figure 3. The FRAM-based estimate of standard deviation (SD) versus the reference value for each of the 8 measurands in data set 1.

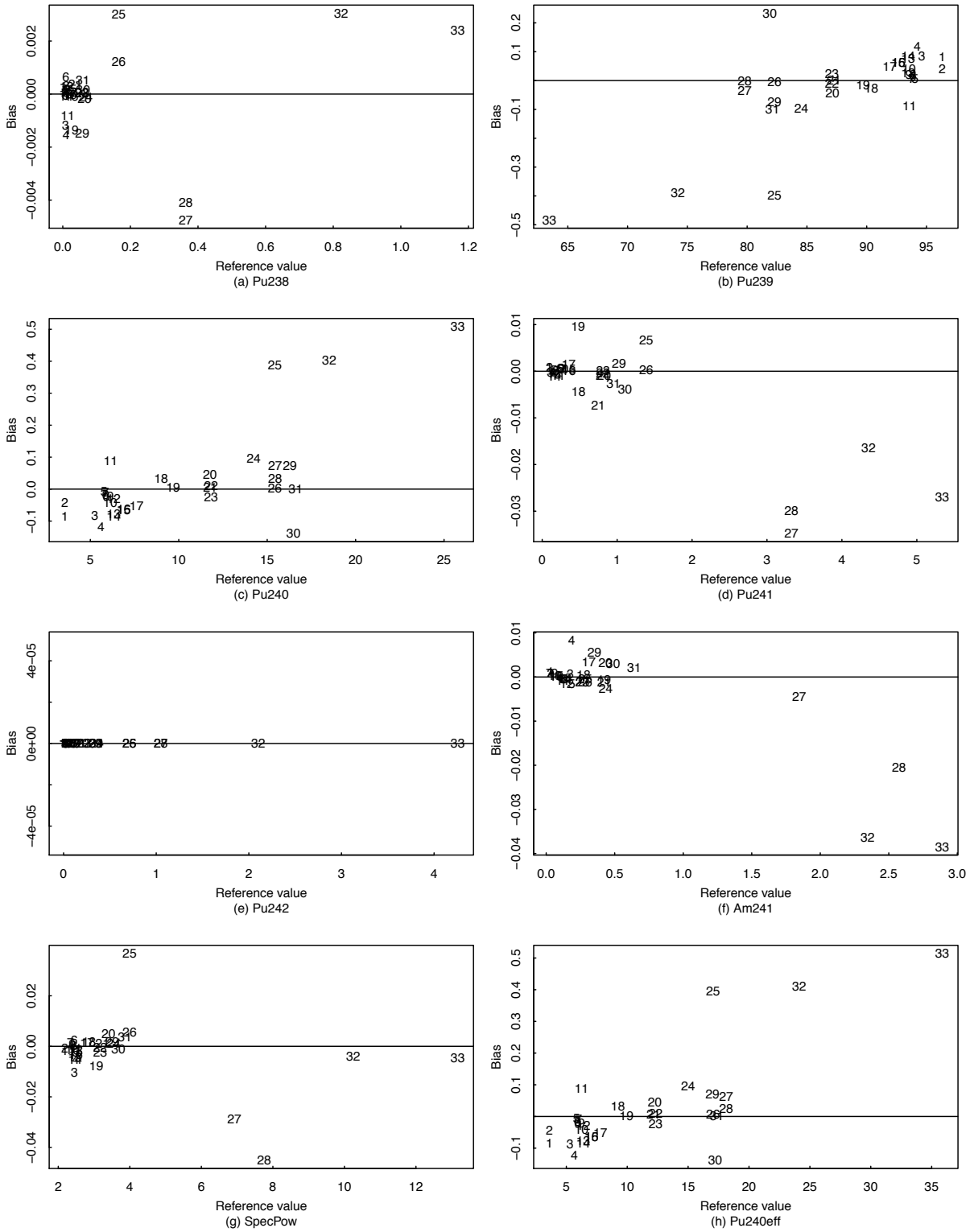

Figure 4. The item-specific bias versus the reference value for each measurand (0 for Pu242); the plotting symbols are the integers 1 to 33, which identify the reference standard in data set 1.

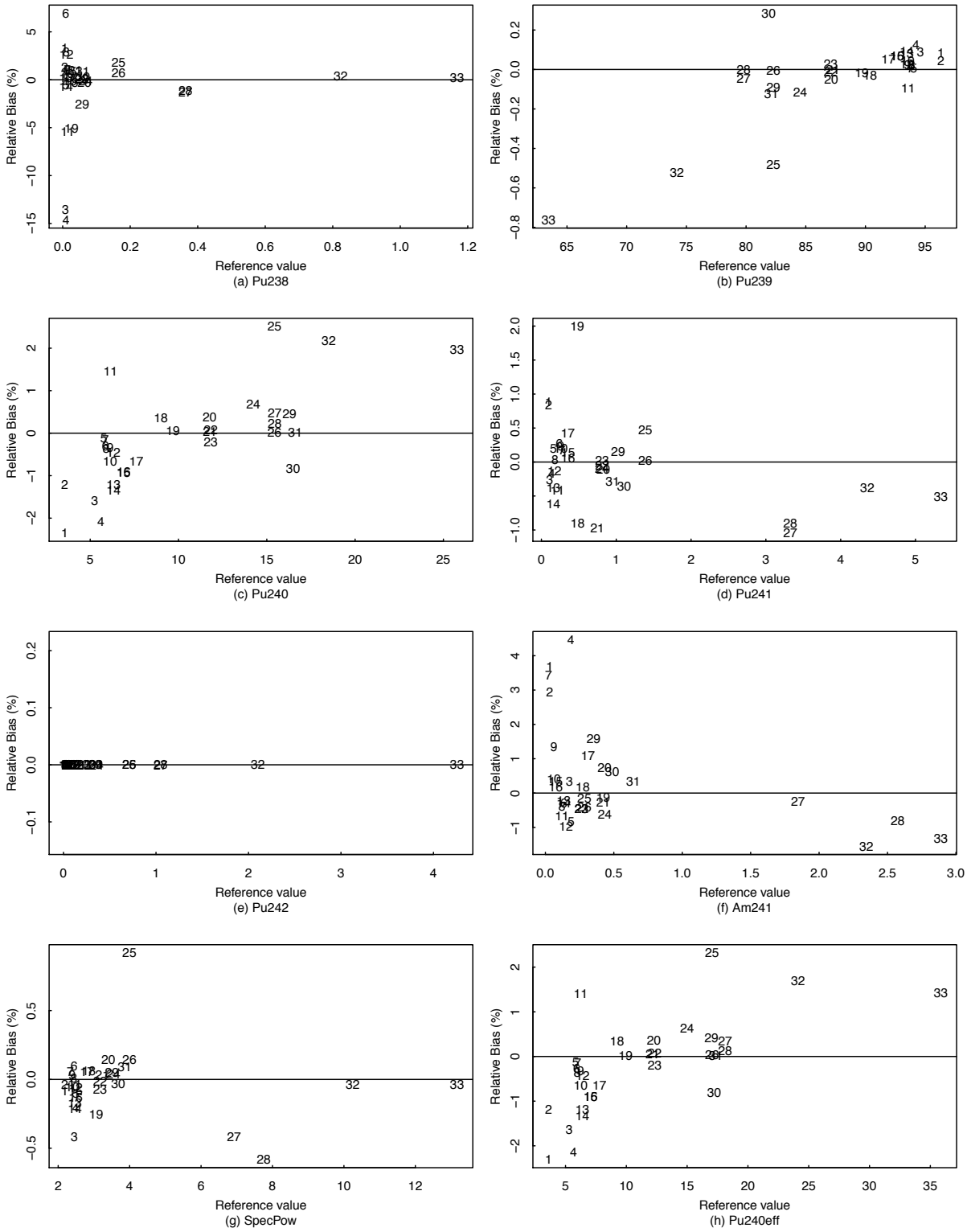

Figure 5. The relative item-specific bias versus reference value (0 for Pu242), and the plotting symbols are the integers 1 to 33, which indicate the identity of the reference standard in data set 1.

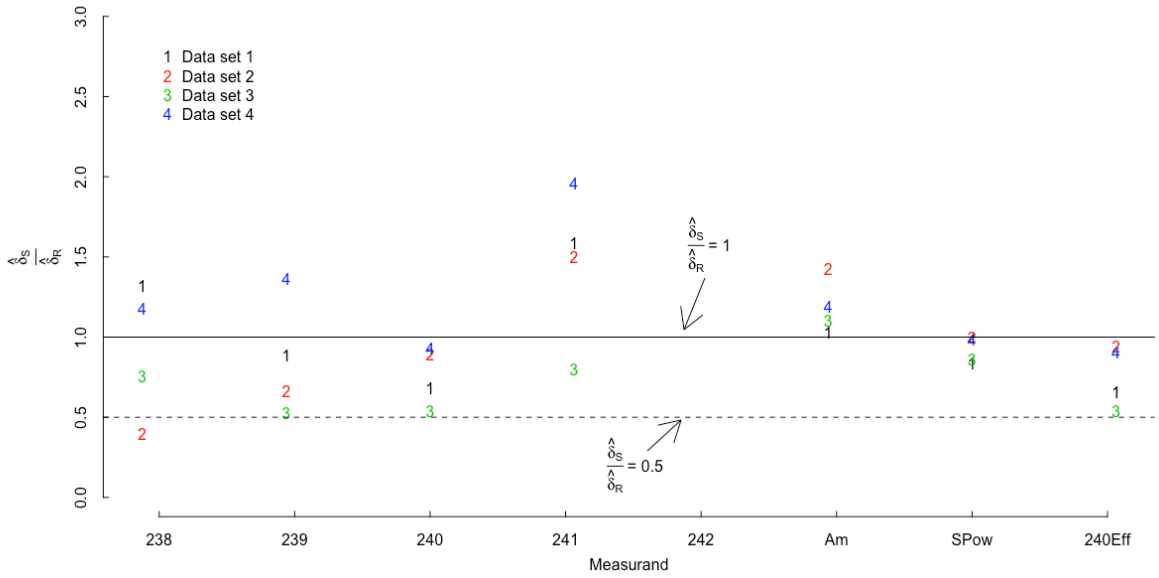

Figure 6. Estimate of the ratio $\frac{\delta_S}{\delta_R}$ for each of the 4 Pu data sets with 8 measurands.

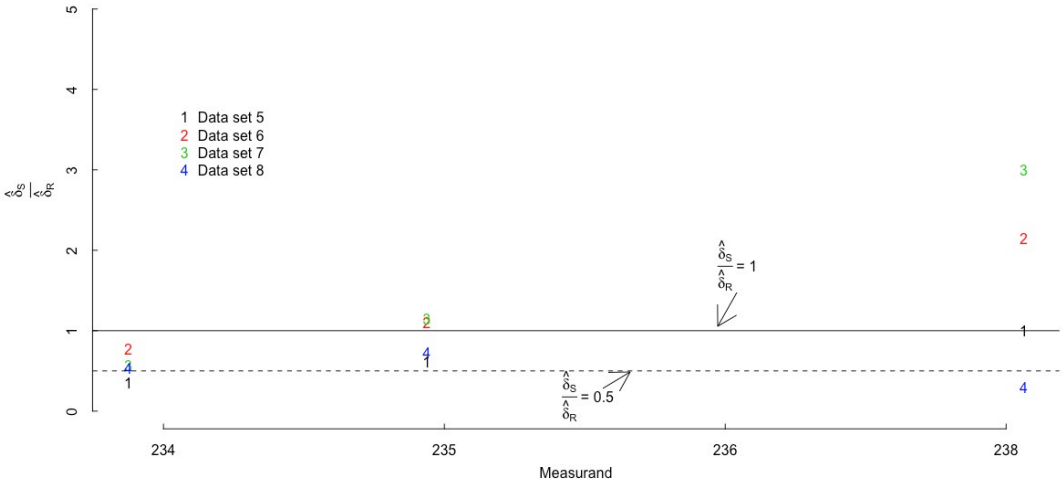

Figure 7. Estimates of the ratio $\frac{\delta_S}{\delta_R}$ for each of the 4 U data sets with 4 measurands.

2. Measurement Error Modelling and Variance Component Estimation

Most of the measurands exhibit larger SD and item-specific bias as the reference value increases, so a reasonable measurement error model is

$$M_{ij} = \mu_i(1 + B + S_i + R_{ij}), \qquad (1)$$

where for measurement j of the ith standard, M_{ij} is the measurement result , μ_i is the true value, B is the long-term bias, S_i is the short-term bias associated with item i, and R_{ij} is a random error. In this

case $X_{ij} = \frac{M_{ij} - T_i}{T_i}$, the relative difference between the FRAM measurement and the nominal value ($T_i \approx \mu_i$) provided by mass spectrometry. The very small measurement errors in mass spectrometry are assumed to be 0 here, resulting in a slight over-estimate of the error RSDs in FRAM. For completeness, Grubbs' estimator was also applied [6], accounting for errors in mass spectrometry; the Grubbs'-based error RSD estimates are very close to those given below using random effects analysis of variance (ANOVA [5]) applied to $X_{ij} = \frac{M_{ij} - T_i}{T_i}$.

In Eq. (1), S_i is often assumed to have a normal distribution with mean 0 and variance denoted $S_i \sim N(0, \delta_S^2)$ and that $R_{ij} \sim N(0, \delta_R^2)$ for all items i = 1, 2., …, g. The normality assumption is not important for variance component estimation, but is important in assessing the uncertainty in the variance component estimates. Regarding notion, the "=" in Eq. (1) should be regarded as "equals in distribution," where the random variable X has the same distribution as a constant $\mu_i(1+B)$ plus the sum of two random variables $\mu_i(S_i + R_{ij})$; and, δ denotes a relative standard deviation while σ denotes an absolute standard deviation. A t-test for overall bias B in Eq. (1) suggests that there is no need for an overall bias term for any measurand (except perhaps for Pu241), so B will be assumed to be zero here and throughout. Item-specific bias is often not negligible, and in that case, a laboratory must expend considerable effort to assess and estimate the variance of item-specific bias, S_i, the short-term bias associated with item i.

Model (1) implies that there is a need to partition error variance into "within" (W) items and "between" (B) as in

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 + n_i\sum_{i=1}^{g}(\bar{x}_i - \bar{x})^2 = SSW + SSB \qquad (2).$$

In the experimental data analysed here, the numbers of measurements per item n_i in Eq. (2) vary from 5 to 25, with an average of approximately 15, with a total of 490 measurements of the 8 measurands for the 33 items.

The within-item replicate variance δ_R^2 in Eq. (1) can be estimated using the observed sample variances as those shown in Figure 2. Because the sample sizes n_i in Eq. (2) vary from 5 to 25, a weighted average should be used. And, δ_S^2 can be estimated using the variance of the estimated item-specific biases in Figure 4. However, the estimated item-specific biases include the effect of random error variance, so that must be subtracted off. More formally, it can be shown that

$MSW = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2}{g(n_i - 1)}$ has expected value δ_R^2 and that $MSB = \frac{\sum_{i=1}^{g} n_i(\bar{x}_i - \bar{x})^2}{g-1}$ has expected value

$\delta_R^2 + \frac{1}{N(g-1)}(N^2 - \sum_{i=1}^{g} n_i^2)\delta_S^2 = \delta_R^2 + C\delta_S^2$, where $N = \sum_{i=1}^{g} n_i$ and $C = \frac{1}{N(g-1)}(N^2 - \sum_{i=1}^{g} n_i^2)$.

Therefore, MSW $= \hat{\delta}_R^2$ is an unbiased estimate of δ_R^2 and $\frac{MSB - MSE}{C} = \hat{\delta}_S^2$ is an unbiased estimate of δ_S^2. And, in standard one-way random effects ANOVA [4-7], if the absolute or relative variances are not constant, then it is straightforward to show that MSW is an unbiased estimate of the average relative variance $\frac{1}{g}\sum_{i=1}^{g}\delta_{Ri}^2$ and $\frac{MSB - MSE}{C}$ is an unbiased estimate of the average relative variance $\frac{1}{g}\sum_{i=1}^{g}\delta_{Si}^2$ [7]. Figure 2 indicates the possibility that some of the SDs are outliers and Figures 4 and 5 indicate the possibility that some of the item-specific biases are outliers. And, it is known, for example, that the reference values (provided by mass spectrometry) for items 3, 4, and 11 (items A1-92, A1-86, and STD117) probably need revision for Pu238, as the corresponding FRAM measurements seem to suggest. Also, the observed sample variances that are used to empirically estimate δ_R^2 are assumed to arise from independent replicate measurements.

Recall that because most of the measurands appear to exhibit multiplicative model behaviour such as in Eq. (1), it is helpful to apply ANOVA scaled data $X_{ij} = \frac{M_{ij} - T_i}{T_i}$, and assume that the reference values are so close to μ_i that uncertainty in the reference values can be ignored. On this basis, the percent relative standard deviations of the item-specific estimated biases for the 8 measurands (recall that measurand 5 of 8 is Pu242, which is given here only for completeness) are 4.12, 0.20, 1.10, 0.59, 0.09, 1.42, 0.24, and 1.02, respectively. And, the averages of the observed relative standard deviations are 4.1 (1.9), 0.20 (0.19), 1.1 (1.1), 0.59 (0.52), 0.09 (0.04), 1.4 (0.98), 0.24 (0.16), 1.0 (1.0), respectively, where the values in parentheses are from a robust version (by trimming outliers toward the median) of the observed standard deviations. The averages of the predicted relative standard deviations are 3.9, 0.18, 1.50, 0.36, 0.09, 1.8, 0.36, and 1.5, respectively.

Using $\frac{MSB - MSE}{C}$ to estimate $\frac{1}{g}\sum_{i=1}^{g}\delta_{Si}^2$, the resulting estimates are 4.0 (1.8) , 0.2 (0.09), 1.1 (1.1), 0.56 (0.48) , 0.09 (0.04) , 1.5 (0.96), 0.23 (0.15), and 0.99 (0.97), respectively, where a robust version (less sensitive to possible outliers) of each estimate is given in parenthesis.

Using MSW to estimate $\frac{1}{g}\sum_{i=1}^{g}\delta_{Ri}^2$ the resulting empirical estimates are 3.3 (3.1), 0.2 (0.2), 1.6 (1.5), 0.34 (0.33), 0 (0), 1.4 (1.4), 0.26 (0.26), 1.5 (1.4), where a robust version of each estimate is given in parenthesis. Using MSW to estimate $\frac{1}{g}\sum_{i=1}^{g}\delta_{Ri}^2$ the resulting estimates based on FRAMs estimated RSDs are 4.1, 0.19, 1.6, 0.36, 0.27, 1.9, 0.37, and 1.5, respectively.

The largest (of 33) RSDs in the standards (for the first 6 of 8 measurands; measurands SpecPow and Pu240eff do not have uncertainties assigned to the standards) are 0.11, 0.0002, 0.003, 0.03, 0.08, and 0.13, respectively, so errors in the nominal values were ignored here.

The bootstrap is one option to estimate the uncertainty in the estimates $\hat{\delta}_R^2$ and $\hat{\delta}_S^2$ [5] in which the within-item measurements and the 33 item-specific biases are sampled with replacement to mimic

repeating the experiment (which would consist of taking 490 measurements of another set of 33 items). For example, for the empirical estimates 3.3, 0.2, 1.6, 0.34, 0, 1.4, 0.26, 1.5 given above to estimate $\frac{1}{g}\sum_{i=1}^{g}\delta_{Ri}^2$, the 95% bootstrap confidence intervals are (2.87, 3.67), (0.19,0.25), (1.39,1.66), (0.31, 0.35), (0,0), (1.22, 1.55),(0.23, 0.27), and (1.31,1.57), respectively. And, for the estimates 4.0, 0.2, 1.1, 0.56, 0.09, 1.5, 0.23, and 0.99 given above for the estimate $\frac{1}{g}\sum_{i=1}^{g}\delta_{Si}^2$, the, the 95% bootstrap confidence intervals are (1.7, 5.8), (0.02, 0.17), (0.53, 1.44), (0.32, 0.73), 0.03,0.1), (0.5,1.7),(0.08,0.26) and (0.51,1.33), respectively.

3.Summary

Section 2 gave estimates $\hat{\delta}_R$ and $\hat{\delta}_S$ for the first (of four) Pu data sets. Appendix 1 lists the estimates $\hat{\delta}_R$ and $\hat{\delta}_S$ for the 8 measurands for the 4 Pu data sets and for the 4 measurands the 4 U data stes. As in [4], there is clear evidence that the true value of δ_S is non-zero, and often is substantial, as in Figures 6 amd 7. Future work will include Bayesian ANOVA [8,9]; however, the sanple sizes are large in all 8 data sets, so most Bayesian analyses will agree closely with the frequentist analyses presented here.

Appendix 2 gives a littlle more detail about FRAM's current UQ approach [10] in the context of two main findings of this paper: item-specific bias is non-negligible, and on average over all 8 data sets, FRAM's estimate of δ_R is approximately 10% larger than the empirical estimate $\hat{\delta}_R$.

References

[1] Sampson, T., Kelley, T., Vo, D., Application Guide to Gamma-Ray Isotope Analysis Using the FRAM Software, https://www.lanl.gov/orgs/n/n1/appnotes/LA-14018-M.pdf, LAUR-1408, 2003.
[2] Guide to the Expression of Uncertainty in Measurement (GUM) , BIPM, JCGM 100:2008.
[3] Burr, T., Croft, S., Jarman. K., Nicholson, A., Norman, C., Walsh, S., Improved Uncertainty Quantification in Non-Destructive Assay for Nonproliferation, Chemometrics, 159, 164-173, 2016.
[4] Burr, T., Sampson, T, Vo, D., Statistical Evaluation of FRAM γ-Ray Isotopic Analysis Data, Applied Radiation and Isotopes 62**,** 931-940, 2005.
[5] Miller, R., Beyond ANOVA, Basics of Applied Statistics, Chapman and Hall, 1998.
[6] Walsh, T., Burr, T., Martin, K., Discussion of the IAEA Error Approach to Producing Variance Estimates for use in Material Balance Evaluation and the International Target Values, and Comparison to Metrological Definitions of Precision, Journal of Nuclear Materials Management, 45(2),4-14, 2017.
[7] Vardeman, S., Wendelberger, J., The Expected Sample Variance of Uncorrelated Random Variables with a Common Mean and Applications in Unbalanced Random Effects Models, www.researchgate.net/publication/241769761, 2003.
[8] Burr, T., Krieger, T., Norman, C., Approximate Bayesian Computation applied to Metrology for Nuclear Safeguards ESARDA bulletin 57, 50-59, 2018.
[9] Solari, F., Liseo, B., Sun., D., Some Remarks on Bayesian Inference for One-Way ANOVA models, Annals of the Institute of Statistical Mathematics 60, 483-498, 2008.
[10] Vo, D. FRAM's Isotopic Uncertainty Analysis, Proceedings of the annual meeting of the Institute of Nuclear Materials Management 2005.
[11] Vo, D., FRAM's Isotopic Uncertainty Analysis Advances in Nuclear Nonproliferation Topical American Nuclear Society Advances in Nuclear Nonproliferation Technology, September 2016.

Appendix 1. Summary of estimated RSDs in % for data sets 1:8.
For the Pu cases: columns 2 to 9 are: 238 239 240 241 242 241 Am Spec Pow Pu240 Eff

For the U cases: columns 2 to 4 are 234 235 236 238

Column 2 : $\hat{\delta}_R$

Column 3 : $\hat{\delta}_S$

Column 4 is the effective random error (repeatability and item-specific bias) RSD ("the bottom line"), $\hat{\delta}_T = \sqrt{\hat{\delta}_R^2 + \hat{\delta}_S^2}$. Note that the estimate $\hat{\delta}_T$ is largest for measurand 1 which is Pu238 (not surprising). Column 5 is the standard deviation of the relative differences $\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$ with $X_{ij} = \frac{M_{ij} - T_i}{T_i}$ calculated using all measurements (the left hand side of Eq. (1)). Column 6 is the root mean squared error $\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i} x_{ij}^{\,2}}$. Appendix 1 results are a high level summary, with ANOVA applied to each full data set with no outliers removed (except for data sets 2 and 5 because the outliers were strong enough to have a large influence on the results), and no exploratory data analysis, Column 4 is the "bottom line" and columns 5 and 6 are given for completeness. Data sets 1-4 are Pu and 5-8 are U. For the four Pu data sets, the 8 results are for the measurands Pu238, Pu239, Pu240, Pu241, Pu242 (operator entered, not independently measured so NA=not available is entered), Am241, Specific Power, and Pu240eff. For the four U data sets, the 4 results are for the measurands U234, U235, and U236 (operator entered, not independently measured), U238.

Data set 1: Estimated RSDs in %. Plutonium, Planar Detector, 120–460 keV Analysis These estimated RSDs include the 3 mild outliers seen in Figures 4 and 6 (due to most likely having a slightly bad nominal value for items for items 3, 4, and 11 (items A1-92, A1-86, and STD117).

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i} x_{ij}^{\,2}}$
1	3.4	4.5	5.6	5.6	5.6
2	0.2	0.2	0.3	0.3	0.3
3	1.6	1.1	1.9	1.9	1.9
4	0.3	0.6	0.7	0.6	0.6
5	NA	NA	NA	NA	NA
6	1.4	1.5	2.1	2.0	2.1
7	0.3	0.2	0.3	0.3	0.3
8	1.5	1.0	1.8	1.8	1.8

Data set 2: Estimated RSDs in %. Plutonium, Coaxial Detector, 120–460 keV Analysis.

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i} x_{ij}^{\,2}}$
1	7.0	2.8	7.6	7.6	7.7

2	0.3	0.2	0.3	0.3	0.3
3	1.9	1.7	2.6	2.6	2.6
4	0.4	0.6	0.7	0.7	1.0
5	NA	NA	NA	NA	NA
6	0.7	1.0	1.2	1.2	1.3
7	0.3	0.3	0.4	0.4	0.4
8	1.8	1.7	2.5	2.5	2.5

Data set 3: Estimated RSDs in %. Plutonium, Coaxial Detector, 180–1010 keV Analysis

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	8.4	6.3	10.5	10.4	10.4
2	0.4	0.2	0.5	0.5	0.5
3	2.8	1.5	3.1	3.1	3.2
4	0.6	0.5	0.8	0.8	0.9
5	NA	NA	NA	NA	NA
6	0.7	0.8	1.0	1.0	1.0
7	0.6	0.4	0.7	0.7	0.7
8	2.6	1.4	3.0	3.0	3.0

Data set 4: Estimated RSDs in %. Plutonium, Planar Detector, 60– keV Analysis

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	3.7	4.4	5.7	5.7	5.8
2	0.1	0.1	0.2	0.2	0.2
3	0.9	0.8	1.2	1.2	1.3
4	0.3	0.5	0.6	0.6	0.6
5	NA	NA	NA	NA	NA
6	1.0	1.2	1.5	1.5	1.5
7	0.2	0.2	0.3	0.3	0.3
8	0.9	0.8	1.2	1.2	1.3

Data set 5: Estimated RSDs in %. Low-Enriched U, Coaxial Detector, 120-1010 keV Analysis. Note the large RSD estimates in U234 (measurand 1). The U234 cannot be reliably measured when the U235 fraction is small, such as 1%. So, a separate ANOVA was performed for U234 without the 0.31% and 0.71% U235 items. The results are in the second set of values in row 1, for the 18 of 23 items for which the U235 fraction is larger than 0.71% (so, for example, 20.4% RSD is reduced to 5.7% RSD).

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	20.4 5.7	10 2.0	22.8 6.1	22.7 6.1	23.4 6.5
2	2.1 1.7	1.3 1.3	2.5 2.2	2.5 2.2	2.7 2.4
3	NA	NA	NA	NA	Na
4	0.3 0.3	0.3 0.3	0.4 0.4	0.4 0,4	0.4 0.4

Data set 6: Estimated RSDs in %. High-Enriched U, Coaxial Detector, 120-1010 keV Analysis, with item 3 omitted because item 3 does not have the correct nominal value.

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	2.6	2.0	3.2	3.2	4.6
2	1.0	1.1	1.5	1.4	1.8
3	NA	NA	NA	NA	NA
4	1.3	2.8	3.1	3.0	3.6

Data set 7: Estimated RSDs in %. Low-Enriched U, Planar Detector, 60-250 keV Analysis

This data is new was not analyzed when the v4.2 analysis was done [4]. As in Data set 5, The U234 cannot be reliably measured when the U235 fraction is small, such as 1%. So, a separate ANOVA was performed for U234 without the 0.31% and 0.71% U235 items. The results are in the second set of values in row 1, for the 7 of 24 items for which the U235 fraction is larger than 0.71% (so, for example, 13.2% RSD is reduced to 3.5% RSD).

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	13.2 3.5	2.8 2.0	13.5 4.0	13.5 4.0	13.5 4.1
2	1.7 0.7	1.2 0.8	2.1 1.0	2.1 1.0	2.1 1.2
3	NA	NA	NA	NA	NaN
4	0.1 0.1	0.2 0.3	0.2 0.3	0.2 0.3	0.3 0.3

Data set 8: Estimated RSDs in %. High-Enriched U, Planar Detector, 60-250 keV Analysis
This data is new was not analyzed when the v4.2 analysis was done [4].

	$\hat{\delta}_R$	$\hat{\delta}_S$	$\hat{\delta}_T$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2}$	$\sqrt{\sum_{i=1}^{g}\sum_{j=1}^{n_i}x_{ij}^{\ 2}}$
1	1.7	0.9	1.9	1.9	2.2

2	1.1	0.80	1.4	1.4	1.6
3	NA	NA	NA	NA	NA
4	17.5	5.2	18.4	18.2	18.3

The total number of repeat measurements in data sets 1:8 are: 490, 729, 826, 296, 282, 499, 472, 301, respectively.

The number of repeats for data set 1 for 33 items are:
15 11 20 20 6 15 20 15 15 15 15 25 5 15 15 15 6 6 15 15 8 15 14 20 14 15 15 15 15 15 20 20 20
The number of repeats for data set 2 (45 items):
20 20 10 10 9 10 10 10 11 30 11 20 21 10 21 11 20 10 10 20 29 10 21 11 10
10 20 21 26 10 20 10 20 26 10 10 10 20 21 21 21 21 10 21 26
The number of repeats for data set 3 (53 items):
20 20 10 10 9 10 10 11 10 11 11 30 11 11 11 20 21 10 21 11 11 20 10 10 20
29 10 21 11 11 10 10 20 21 26 10 20 10 20 26 10 11 10 10 20 21 21 21 21 10
21 20 26
The number of repeats for data set 4 (23 items):
10 14 14 10 10 10 14 10 14 10 14 10 14 8 24 15 14 8 9 15 21 15 13
The number of repeats for data set 5 (22 items):
8 8 22 14 15 15 8 8 8 8 12 13 15 21 15 9 8 14 15 24 8 14
The number of repeats for data set 6 (34 items):
15 15 11 20 20 6 15 20 15 15 15 15 25 5 15 15 15 6 6 15 15 8 15 14 14 14 15 15 15 15 15 20 20 20
The number of repeats for data set 7 (24 items):
19 20 17 19 15 17 27 16 18 29 16 14 26 16 13 21 23 21 28 23 21 20 17 16
The number of repeats for data set 8 (16 items):
17 20 21 23 19 17 16 19 16 20 20 16 12 23 12 30

The t-test for overall bias:
```
> t.test(means)
t = -1.9227, df = 32, p-value = 0.06345
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.0066651433  0.0001922504
sample estimates:
   mean of x
-0.003236446
# result: no overall bias
```

To generate Tables 1,2,3,6 for the 8 individual measurands (and similarly for Tables 4,5,7,8):
```
for(i in 1:8) {
dtemp <- ts1d[,i]
cat(c(i,round(estvars(groups=temp.indices,d=dtemp)[1:5]^.5*100,3)),"\n")
}
```

1 3.384 4.491 5.624 5.568 5.621
2 0.225 0.209 0.307 0.305 0.309
3 1.581 1.081 1.915 1.905 1.907
4 0.341 0.557 0.653 0.646 0.646
5 0 0.085 0.085 0.084 0.089
6 1.432 1.486 2.063 2.047 2.077
7 0.261 0.227 0.346 0.343 0.346

8 1.49 0.986 1.786 1.778 1.784

The R function estvars0 implements the one-way ANOVA described in Section 2; it gives essentially the same results as random effects ANOVA in R using lmer() as shown next.

```
estvars = function(groups,d) { # groups vector for group memberships; d for relative differences
N = length(d); ngroups =length(unique(groups))
vard.vec = numeric(ngroups); meand.vec = numeric(ngroups); n.vec = numeric(ngroups)
temp.indices = as.numeric(names(table(groups)))
for(i in 1:ngroups) {
indices = (1:N)[groups==temp.indices[i]]
n.vec[i] =length(indices)
dtemp =d[indices]
vard.vec[i] = var(dtemp)
meand.vec[i] = mean(dtemp)
}
dmean = mean(d)
temp1 = sum(n.vec*(meand.vec-dmean)^2)
temp2 = sum((n.vec-1)*vard.vec)/sum(n.vec-1)
temp3 =N*(ngroups-1)*(temp1/(ngroups-1) - temp2)/(N^2 - sum(n.vec^2))
c(temp2,temp3,temp2+temp3,var(d),mean((d)^2))
}
```

If the group sample sizes are equal, estvars() results agree with lmer() results. If not, then there are very small, negligible differences between estvars() and lmer() results. The lmer() function implements restricted maximum likelihood [6], while estvars() is a method of moments estimator.

```
tempgroups = c(1,1,1,1,2,2,2,2,3,3,3,3)
#tempd =  c(rnorm(n=4),rnorm(mean=1,n=4),rnorm(mean=2,n=6))
# tempd rounded:
tempd = c(-0.40,  1.22, -0.32, -1.38,  0.49,  0.72,  1.69,  1.47,  1.15,   4.31,  2.00,  1.46)
temp = lmer(tempd ~ 1 + (1 |tempgroups))
summary(temp)
Random effects:
 Groups     Name        Variance Std.Dev.
 tempgroups (Intercept)  1.209    1.099
 Residual                1.178    1.085
> round(estvars0(groups=tempgroups,d=tempd),3)
[1] 1.178 1.209 2.387 2.057 2.955. # the 1.209 and 1.178 agree with lmer() results.
```

Appendix 2. Version 4.2 versus version 5.2 of FRAM; Uncertainties in the Reference items, and Possible Reasons for Item-specific Bias; FRAMs Bottom-up Variance Estimation

Appendix 2 addresses three main topics: (1) version 4.2 versus version 5.2 of FRAM, (2) uncertainties in the reference items, (3) possible reasons for item-specific bias, and (4) error variance estimation provided in version 5.2.

First, the RSDs in this report for v5.2 of FRAM are acceptably small, but slightly larger than those reported in [4] for version 4.2 of FRAM. One feature that is different between v4.2 AVOVA data and V5.2 data is how the tailing constants are treated. In v4.2 analysis the tail constants were fixed while they are free in 5.2; this might contribute to the increased variability observed in v5.2 (this report).

Second, Table A2.1 provides notes on uncertainties assigned to the mass spectometry values for FRAM v5 ANOVA for data set 1.

Sample Name	Comment on Uncertainty
CBNM93, CBNM84, CBNM70, CBNM61	Certified Reference Material from Geel. Uncertainty listed is ½ of the 95% confidence interval listed on the certificate. No corrections made for decay. Circa mid 1980s.
STD117, STD118, STD119, STD120, STD116	One LANL mass spec from 1979. Uncertainty reported as "Deviation" and it is not clear whether it is supposed to approximate a 1 or 2 sigma value. Will report it as if it was 2 sigma. Uncertainty for this analysis is "Deviation"/2.
CALEX	Working Reference Material used in the DOE Calorimetry Exchange Program (now inactive). Isotopic composition reevaluated by NBL in 2008. Uncertainties are ½ of their estimated 95% confidence limit. No corrections made for decay.
A1-92	One mass spec from ORNL in 1970. No original mass spec uncertainty information available. I arbitrarily assign the same uncertainties as STD117 because its values are old (1979) and the isotopic distribution is similar.
A1-86	No original mass spec or mass spec uncertainty information available. Same uncertainty assignment as for A1-92 above.
LAO225	Three analyses at two LANL labs circa 1983. Uncertainty is 1 std dev of the mean of the three sets of analyses.
STD40	Averages of analyses on 4 samples at two LANL facilities in 1980-1982 time frame. Uncertainty is the std dev of the mean of all the analyses.
SRPISO3,6,9,12,15	Analyses from 3 labs. Uncertainty is the standard deviation of the mean of the three lab results. There are no systematic uncertainties included such as uncertainty contribution from standards
STD121	Same material as LAO225.
PUEU7	No uncertainty information has survived. Uncertainty estimated as the standard deviation of a single measurement of the SRPISO6 sample.
STDR3, JOO1325, STD8, STD6, STD3	Assume same uncertainty as assigned for STD117. The items are from the same era (late 1970s –mid 1980s and probably have the same quality routine single LANL mass spec measurement.
LAO256	Same vintage and program as LAO225. Use the single measurement uncertainty from LAO225. Double the uncertainty for Am.

Table A2.1. Notes on uncertainties assigned to standards. None of the uncertainty assignments are decay corrected. Most of the uncertainties do NOT include systematic uncertainty contributions from mass spectrometry calibration. Only the CBNMs and CALEX come from recognized standards organizations and have statements of uncertainty supplied. Others come from recognized mass spec labs and are reliable measurements but the only uncertainty data available is that from repeated measurements or the averages and standard deviations of measurements from several labs. Because no systematic uncertainties were supplied, care has been taken not to place too much weight on averages from repeated measurements from labs that lead to an unrealistically small estimate of the measurement uncertainty.

Third, as in v4.2 analysis in [4], there is evidence that the RSD of item-specific bias is approximately the same on average as the RSD of within-item repeated measurements. Therefore, it is useful to try to gain a better understanding of what factors contribute to item-specific bias. As an example of many possible sources of item-specific bias, there are background difficulties in the 160 keV region, partly due to the backscatter peak in this region. For all items, accurate peak areas are one key to a bias-free analysis. The most crucial part of obtaining an accurate peak area is the ability to accurately subtract the background continuum from beneath the peak being analyzed. This is relatively easy for clean, single peaks on a smoothly varying background. However, if there are complex closely-spaced, overlapping peaks, it often becomes difficult to accurately define the background continuum beneath the complex because with a wide complex of peaks, spanning many keV in energy, the background continuum underneath the complex may have hidden structure or change shape, effects that are not usually visible to the naked eye.

As an example, analysis of Pu240 in the Plutonium, Coaxial detector, 120-460 keV region comes solely from a single peak at 160.3keV (Figure A.1). Note that there is much better resolution of this peak from a planar detector, leading to better peak separation, The plutonium spectra in this region taken with a coaxial detector usually show a global curvature. A feature of the background continuum in this region is that the slope of the background continuum changes under the 160 keV peak complex. There is a physical reason for this artifact—backscatter peak edges from prominent higher energy peaks in the plutonium spectrum. So, as the nature of the item changes, mass, isotopic composition, scattering geometry, etc., the nature of the background continuum in this region changes. For example, compared Figure A.2 (a zoom near the 160keV region of Figure A.1) to Figures A.3 (the same as Figure A.2 but for an item having higher Am241 ppm (726 ppm) and A.4 (the same as Figure A.2 but for an item with 6.30 % Pu240, 240 g Pu as PuO2, , 1757 ppm Am241).

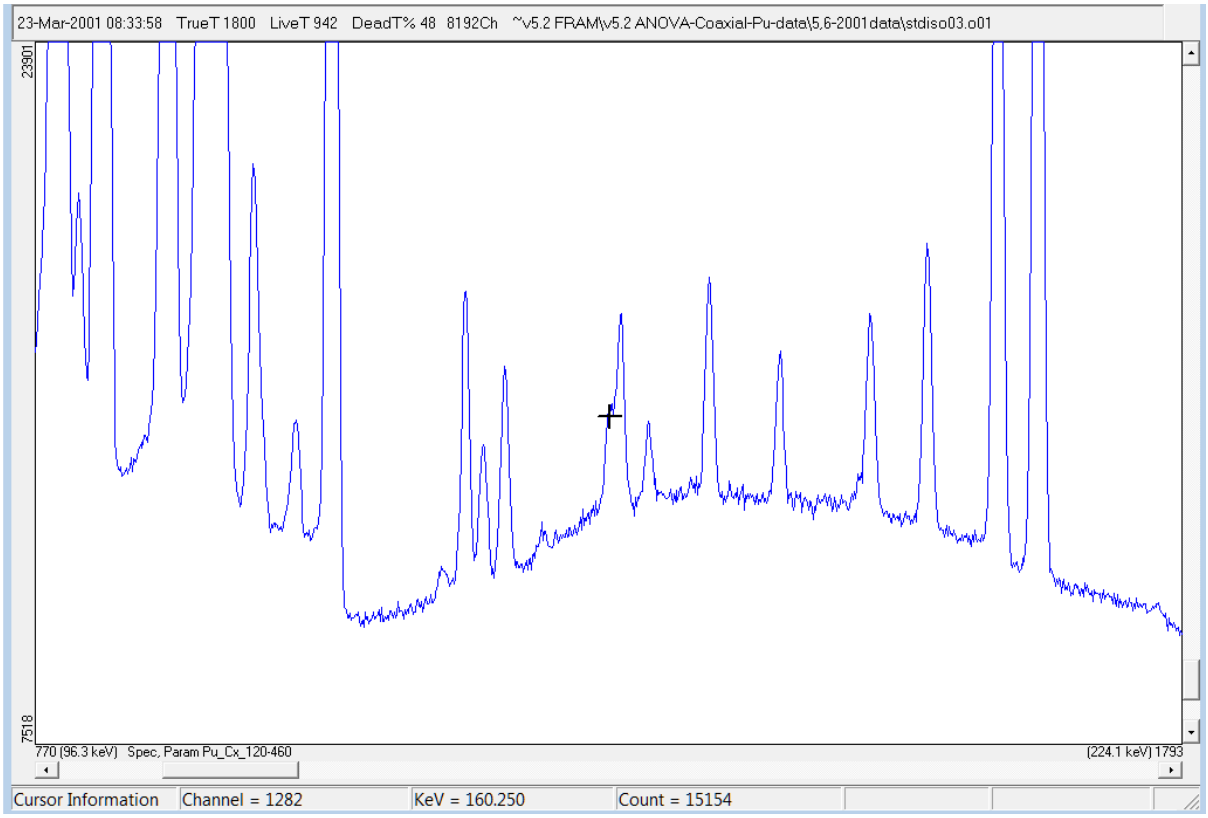

Figure A.1. Global curvature around 160 keV. 3.56% Pu240

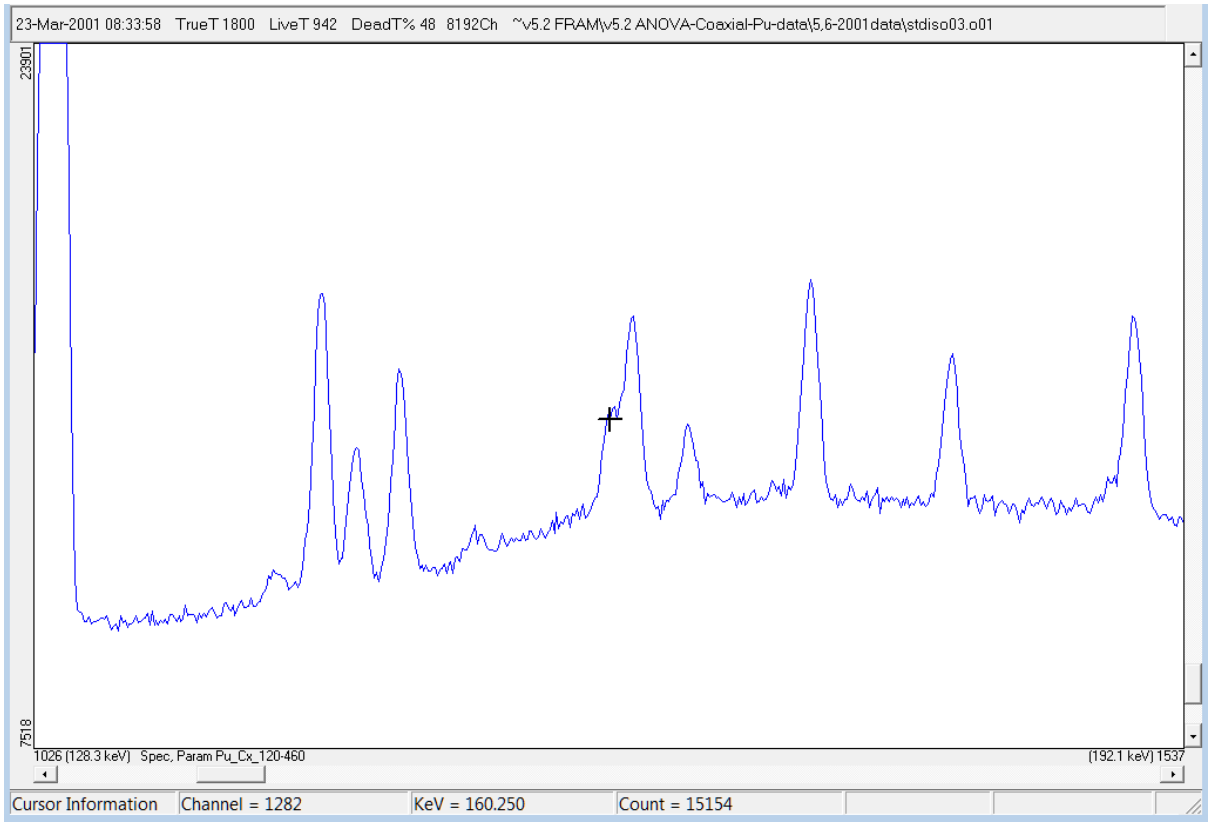

Figure A.2 A zoom to the region near 160keV in Figure A1.1 for . 3.56% Pu240.

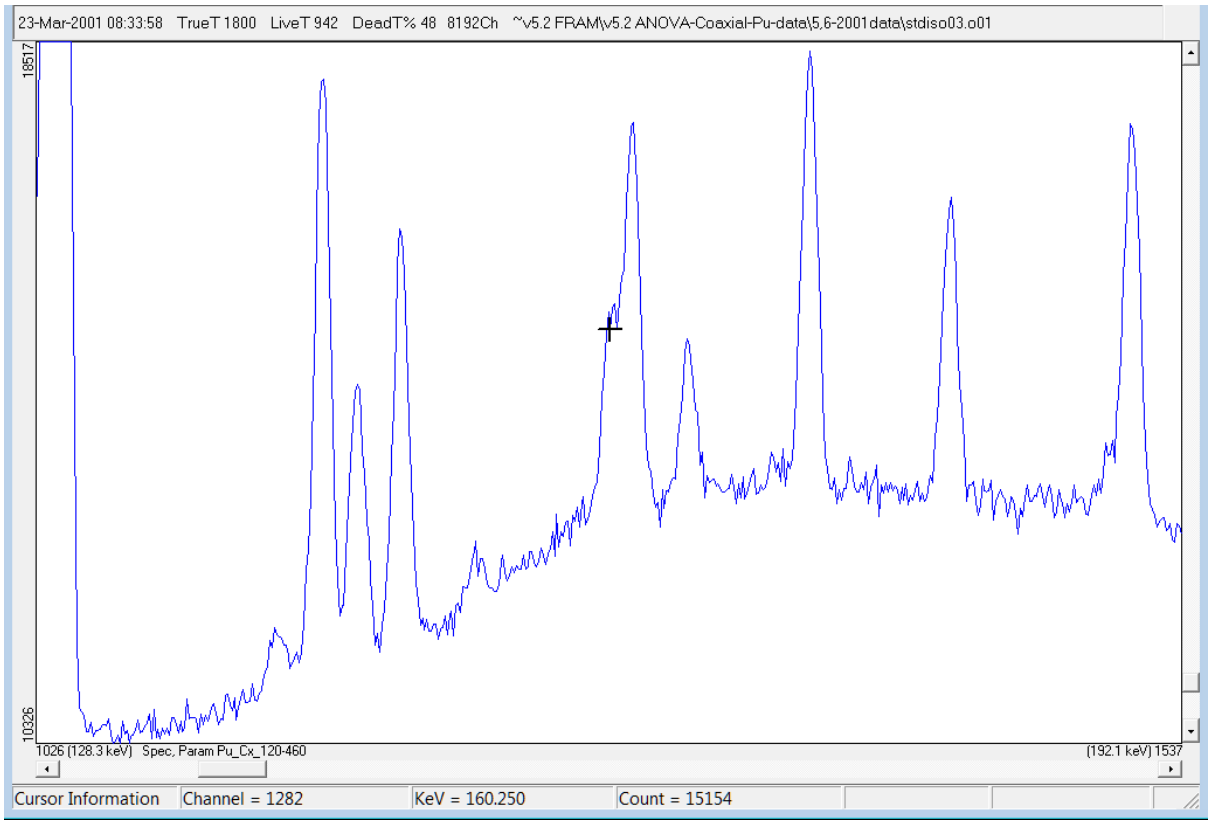

Figure A.3. The same as Figure A.2, but for 726 ppm Am241 with the 3.56% Pu240, leading to a change in the slope of the background compared to Figure A.2.

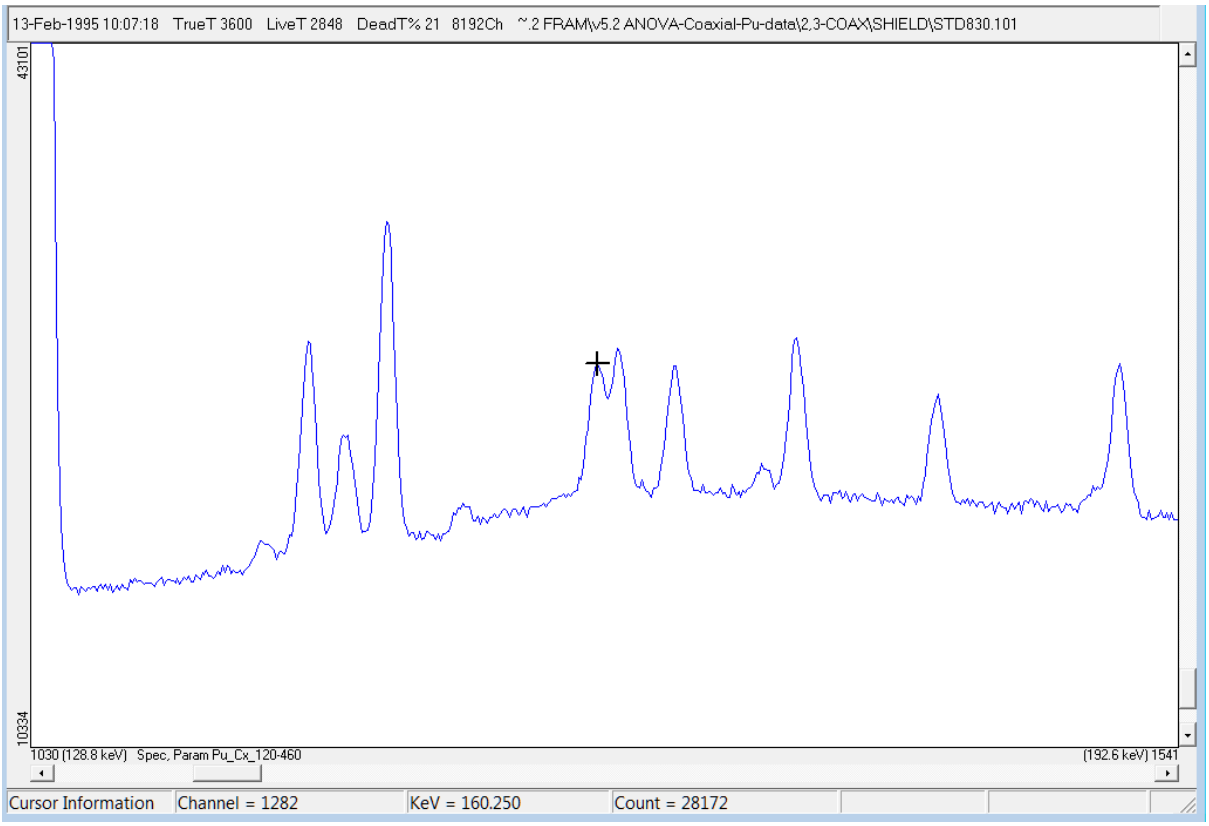

Figure A.4. The same as Figure A.2, but for 1757 ppm Am241 with the 6.30% Pu240, leading to a change in the slope of the background compared to Figure A.2.

In FRAM the user has a choice of 7 different functional forms to define the background continuum underneath any peak region. Typical a simple linear function is chosen in this region, but the concave upward shape of the global continuum in this region generally means that the net peak area results are biased high. This type of changing background continuum shape can have an impact when a single set of parameter file constants are applied to a wide range of samples.

A change in the definition of the regions defining the linear background near 160 keV that is likely to be a main cause of the difference between the results in FRAM v4 versus the results in FRAM v5.

Fourth, recall that FRAM's estimated RSD is slightly larger on average than the observed within-item RSD (by approximately 10%). The within-item error variance estimation provided in version 5.2 is described in [10,11].